

# Herramientas digitales 2019-20

## Módulo: Texto digital

Profesor: Nicolás Vaughan  
(n.vaughan@uniandes.edu.co)

### Tarea No. 1: Markdown, RegEx y Pandoc

**Fecha de entrega:** a más tardar el viernes 30 de agosto de 2019, 4:00 p.m.

**Modo de entrega:** Envíen a mi correo los dos archivos resultantes (Markdown y HTML), con su apellido como nombre de archivo (por ejemplo, vaughan.md, vaughan.html). El asunto del correo debe ser HD Tarea No. 1.

## 1 Descripción corta

Deben procesar un archivo de texto plano, transformándolo con RegEx en uno de formato Markdown, y finalmente en uno de formato HTML, con ayuda de Pandoc.

## 2 Descripción detallada

Usando el motor RegEx del editor de texto Atom, deben procesar el texto *La gente cursi: Novela de Costumbres Ridículas*, de Ramón Ortega y Frías (disponible aquí: <http://bit.ly/2ZpPBEz>). Deben guardarlo como un documento de Markdown (con extensión .md). (Eliminen las primeras 62 líneas, y la parte en inglés al final, que no hacen parte del texto.)

Procesen el texto con RegEx (buscando y reemplazando) para hacer lo siguiente:

1. Unan las líneas (eliminando los saltos de línea) que componen los párrafos, teniendo cuidado de no eliminar los párrafos también. (Como vimos en la sesión virtual.)
2. Los dobles guiones (–) conviértanlos en rayas (—).<sup>1</sup>
3. Los comillones («») conviértanlos en comillas dobles (“”).<sup>2</sup>

---

<sup>1</sup>Las rayas son los guiones largos que se usan, entre otras, para indicar diálogos. Cf. <http://bit.ly/2MB0i19>

<sup>2</sup>Deben ser comillas dos de apertura y cierre (“”) y no comillas dobles rectas ("").

4. Eliminen los espacios en blanco dobles (triples, etc.) entre palabras y al principio de las líneas.
5. Reemplacen por asteriscos los guiones bajos que encierran algunas palabras (por ejemplo, de: `_cursi_` a: `*cursi*`).
6. Dejen los títulos de capítulos en *una sola línea*. Por ejemplo, de:

CAPÍTULO I

La mujer casamentera.

pasen a:

CAPÍTULO I: La mujer casamentera.

Noten los dos puntos (:) luego del número romano.<sup>3</sup>

7. Usando la sintaxis de Markdown, marquen los capítulos como encabezados (*headers*) de nivel 2, como vimos en clase. (No se olviden del Epílogo y el Índice al final.) Antes de cada capítulo pongan una línea de separación (usando `***` en Markdown).
8. Marquen el título del libro y el autor como encabezados de nivel 1.
9. Limpiesen el índice del final del documento y denle formato de lista no numerada en Markdown.
10. Pongan los siguientes nombres en **negrita**: “Robustiana”, “Juanito”, “Alfredo”, “Bonacha”, “Pascual”, “Paquita” y “Adela”. (Esto se puede hacer esto en un solo paso, usando un grupo de captura alternado (`Alfredo|Bonacha|...`)).
11. Finalmente, conviertan el documento a HTML usando Pandoc.

---

A continuación encontrarán el texto resultante en PDF para que vean cómo debe quedar. (Lo hice abriendo en Chrome el HTML y luego imprimiéndolo como PDF.)

---

<sup>3</sup>Para este paso deberán usar grupos de captura (*capture groups*) de RegEx, para capturar todo el texto desde “Capítulo” hasta el final de la línea. (Usen paréntesis para eso en el campo de búsqueda, y luego `$1` en el campo de reemplazar. Vean <http://bit.ly/322jFDI>)